AD_____

Award Number: W81XWH-11-1-0487

TITLE: Development of Advanced Technologies for Complete Genomic and Proteomic Characterization of Quantized Human Tumor Cells

PRINCIPAL INVESTIGATOR: Dr. Robert Moritz

CONTRACTING ORGANIZATION: Institute for Systems Biology
Seattle, WA 98109-5234

REPORT DATE: July 2013

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
              Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                        Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2013	Annual	15 June 2012 - 14 June 2013

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Development of Advanced Technologies for Complete Genomic and Proteomic Characterization of Quantized Human Tumors	W81XWH-11-1-0487
	5b. GRANT NUMBER W81XWH-11-1-0487
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)	5d. PROJECT NUMBER
Dr. Robert Moritz	
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
E-Mail: Robert.Moritz@systemsbiology.org	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Institute for Systems Biology Seattle, WA 98109-5234	

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

With the establishment of GBM cell lines and quantized cell populations form these GBM patients tumor samples we are able to complete some of our aims of our project. We will continue to collect tumor samples with consent from families of GBM patients in preparation to perform the molecular analysis of these. Our efforts in the development of quantized cells and genomic technology to interpret specific transcriptome signatures from quantized cell populations has been completed and is now applied to several GBM cell lines and their quantized sub-populations. Whole genome sequencing from families from our group is now well established and has been published.. Our efforts in the development of quantitative assays for essentially every human protein have been successfully established. We have established cell culture conditions to enable proteomic analysis of quantized cell sub populations. This program will be will provide new technical capability that has already provide significant results and will be completed over the next 12 months.

15. SUBJECT TERMS

Human cohorts, Glioblastoma, Genomic, Proteomic, Single-cell technologies, Hypothesis-driven, integrative systems approach, Early diagnosis,Patient stratification,Blood protein biomarkers,Quantized cell populations

16. SECURITY CLASSIFICATION OF: U			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU		**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

INTRODUCTION

We will develop clinically quantitative tools for direct application to human cohorts with glioblastoma classified cancer. This program promises to deliver important insights to cancer mechanisms (disease-perturbed networks), as well as blood biomarkers to assess progression and stratification of human glioblastoma. This proposal will significantly advance genomic, proteomic and single-cell technologies, enabling the commencement of hypothesis-driven integrative systems approaches to disease (cancer). To this end, we are developing new strategies for advanced genome sequencing, new technologies for the analyses of transcriptomes, miRNAomes and single cells as well as multiplexed quantitative protein measurements including the measurement of isoforms, and post-translational modifications. The tools proposed here will be generally applicable to all cancer-based studies, as the nature of the tool development is designed to identify and quantify DNA, RNAs, proteins and cells, challenges ubiquitous to all human disease systems.

To complete these tasks, we will use a logical approach with the following aims:
Specific Aim 1. Isolate up to 1000 cells from each of five human glioblastomas and quantify initially 500 different transcripts from each cell (transcription factors, CD molecules, relevant signal transduction pathways, etc). Determine whether computational analyses can classify these cells into discrete quantized cell types.
Specific Aim 2. Sort the disassociated tumor cells from several glioblastomas into their quantized cell populations using cell sorting/CD antibodies to each quantized cell type for functional analyses and establish primary cell lines. These will be used for molecular analyses— at the genome, transcriptome, miRNAome and selected proteome levels.
Specific Aim 3. Assess 20-40 candidate blood biomarkers in the bloods of 100 glioblastoma patients with regard to their ability to stratify disease, assess disease progression and predict at an early stage the reoccurrence of the glioblastoma (early detection). Eventually we will use these biomarkers to assess the effectiveness of therapy.
Specific Aim 4. Ten to 20 cells from each major quantized glioblastoma cell type from two patients will be used to determine the complete genome sequences. We will also determine the normal genome sequences of each patient and their family members to enable the Mendelian-based error correction process recently described in our recently published Science paper (*1*). The mutations will be analyzed against quantitative changes in the transcriptomes, miRNAomes and proteomes and against the relevant biological networks.
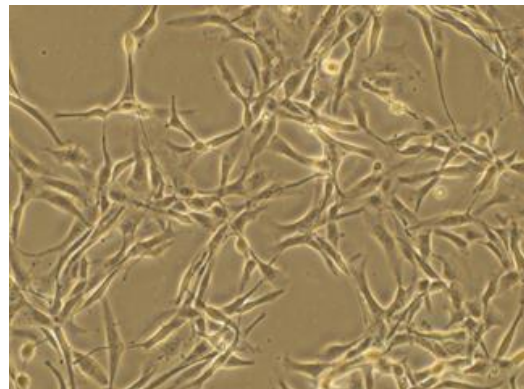**Specific Aim 5**. Analyze the quantized cell populations for their responses (transcriptome, miRNAome, etc) to the perturbations of key glioblastoma-relevant molecules (e.g. nodal points in networks) by RNAi perturbations as well as their responses to glioblastoma-relevant drugs and natural ligands. These assays will be carried out in the laboratory of our collaborator Dr. Greg Foltz at Swedish Hospital.

The expected outcomes and deliverables of this innovative program will be: **1**) deeper understanding of human glioblastoma disease mechanisms; **2**) blood protein biomarkers for use in early diagnosis, stratification of glioblastomas, assessment of the progression of a glioblastoma, assessment of effectiveness of drug treatment and detection of reoccurrence at an early stage; **3**) new strategies for genomic sequencing of quantized cancer cells and their normal counterparts to identify cancer-driver mutations; **4**) new technologies for transcriptome, miRNAome, proteome and single-cell analyses, and **5**) the creation of quantized glioblastoma cell lines that can be used for general molecular characterization as well as to assess the biology of this cancer (drugs, RNAi's, natural ligands) and the effectiveness of existing drugs in reacting with these cell types.
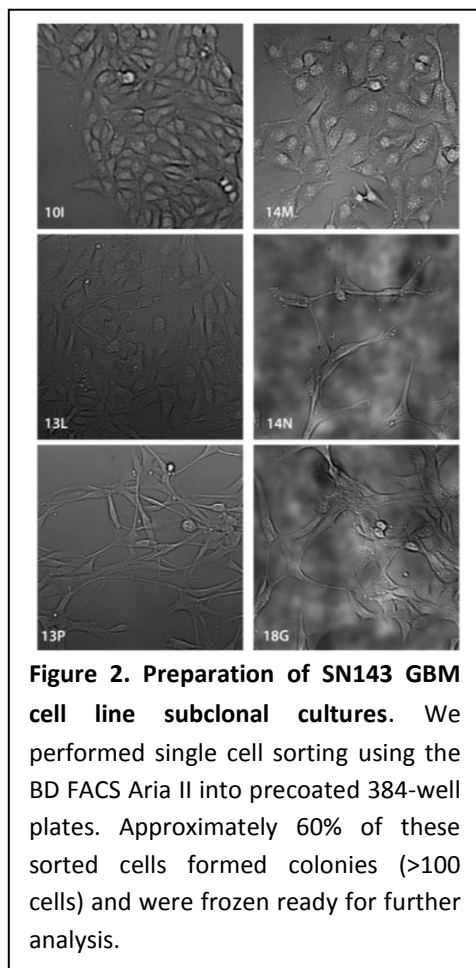
BODY

Aim 1 & 2. Our work is linked with the Ivy Center for Advanced Brain Tumor Treatment at the Swedish Neuroscience Institute (SNI) collaborative group (CA100459P1, Award Number W81XWH-11-1-0488, Swedish Health Services, Dr. Gregory Foltz) to provide cells from human glioblastoma tumors (GBM) excised during surgery at the Swedish medical Center by Dr. Gregory Foltz. During the past year, we have continued to collect GBM samples for our aim 1 to expand the patient cohort available for molecular analysis, genome sequencing, and quantitative assays. To date, the SNI has collected tumor tissue eligible for this program from over forty GBM patients. We have made excellent progress in the establishment of primary GBM cell lines from patients undergoing tumor resection at SNI. The stem cell phenotype has

been confirmed in these cultures by functional assays of self-renewal, differentiation potential, and tumor propagation in vivo where average tumor volume (n = 5) in immuno-compromised mice six weeks after implantation of GBM-patient derived cells increase by average 5-fold. Several of these GBM-derived cultures have been transferred to our group for the generation of quantized cell populations for molecular analyses at the genome, transcriptome, miRNAome and proteome levels (Specific Aims 1 & 2). For each of these established cell lines, a number of single cell clones have been successfully established from the corresponding parental cell lines.

Figure 1. Establishment of SN143 GBM cell line. Culture protocol used from Pollard SM et al., 2009, Cell Stem Cell. Briefly, these cells were cultured on plates coated with laminin and grown under Serum-free conditions with stem cell media and addition of B27 and N2 supplement, growth factors: EGF and bFGF. The doubling time is ~3-5 days.

The second aim of the project is to establish quantized cell populations from the primary

Figure 2. Preparation of SN143 GBM cell line subclonal cultures. We performed single cell sorting using the BD FACS Aria II into precoated 384-well plates. Approximately 60% of these sorted cells formed colonies (>100 cells) and were frozen ready for further analysis.

glioblastoma tumors, so that we can conduct whole genome sequencing (WGS), transcriptomics, and proteomics analyses of these quantized cell populations for novel biomarker discovery. The challenge for this aim is that we will need, as control, blood cells from both patients themselves and their family members. As expected, recruiting both patients and family members proved to be a difficult task. Fortunately, our clinical collaborators at Swedish Hospital managed to complete recruitment of the first patient family with all specimens required in June 2013. In anticipation of this delay, we had decided to proceed with two cell lines for which quantized cell populations have been established (SN143 and SN186, see Fig. 1) and have been developing single cell clonal culture techniques by using these previously collected glioblastoma tumor samples. We have adopted a published protocol for culturing tumor cells in adherent conditions. We modified the protocol by integrating single cell sorting using the BD FACS Aria II into precoated 384-well plates. Approximately 60% of these sorted cells formed colonies (>100 cells) and were collected and froz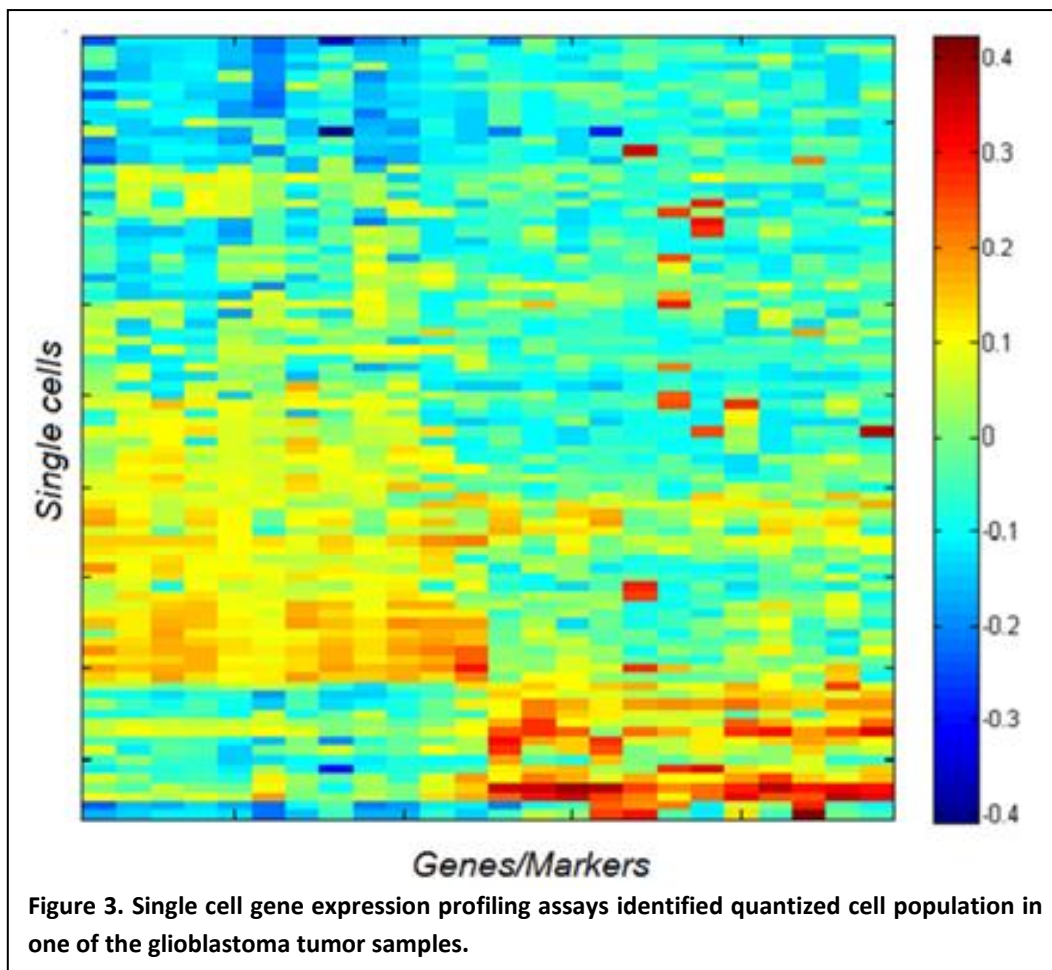en ready for further analysis (Fig 2). As shown in Figure 2, we have established dozens of clonal cultures from 2 patient samples. These clonal cultures exhibit distinct morphological phenotypes, and each clone presumably carries a uniform genome, thus is ideal for subsequent WGS analysis.

We proposed to assess this through transcriptomic profiling of hundreds of single cells. Expression patterns of the selected genes among these single tumor cells would digitally stratify tumor cells into distinct populations. As proof-of-principle, we characterized ~200 cells from GBM patient tumor sample SN143. We conducted single cell gene expression profiling by employing two technological platforms: the Fluidigm Biomark and the NanoString nCounter. Transcriptomic analysis via DNA microarray of six subclones unveiled three quantized cell populations. We derived a panel of 48 genes based on the gene expression profiles from the microarray data and purchased Taqman assays for real-time PCR experiment on the Fluidigm BioMark System. As a comparison with competing technology, we have also custom designed probe sets for the same 48 genes to be used for the Nanostring nCounter platform.

As shown in Figure 3, by analyzing ~100-200 single cells using a panel of 48 markers, we were able to stratify the primary glioblastoma cells into at least three distinct populations. We have
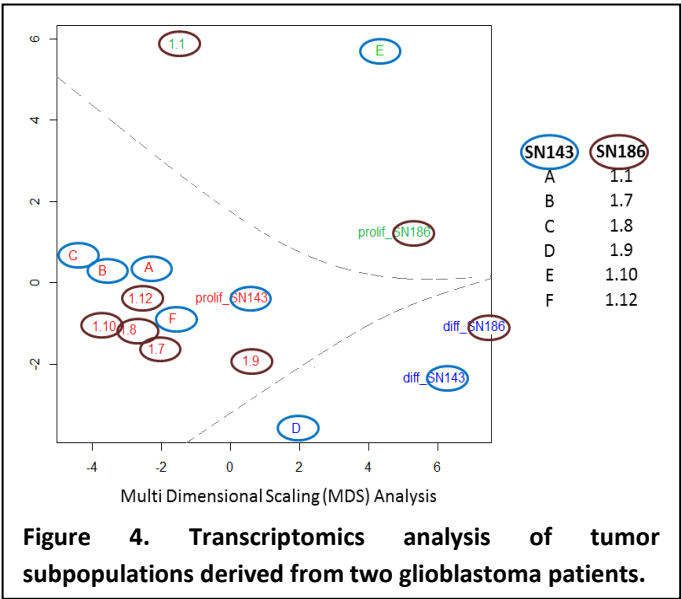
also created quantized cell populations and analyzed the second patient tumor sample (SN186), and observed similar distribution patterns of the cell populations. Thus, we have demonstrated, in principle, that there are indeed quantized cell populations in primary glioblastoma cells, and that single cell gene expression assay is capable of unveiling intratumoral cellular heterogeneity.

To evaluate the molecular heterogeneity of these clonal cultures at the transcript level, we conducted whole transcriptomics analysis on selected clones derived from the first two patient tumor samples (SN143 and SN186). We chose eight cultures from each patient, including the parental cells, differentiated cells, and six single clones, for the gene expression analysis. We then conducted multi-dimensional scaling (MDS) analysis based on the gene expression data. As shown in Figure 4, we were able to identified distinct clusters of these clonal cultures from the two patient samples being analyzed. This cell population distribution pattern is consistent with that observed in our single cell gene expression analysis (figure 3).

Figure 3. Single cell gene expression profiling assays identified quantized cell population in one of the glioblastoma tumor samples.

We further investigated the colony-forming capability of different clonal cultures from one of the patient samples. As shown in Figure 5, clone C forms the largest colony among all the subclones, suggesting that this population may contain more tumorigenic cells.

To further analyze these differences, we chose two genes from our panel of 48 target genes that exhibited the lowest expression in our previous round of single cell analysis, so that to minimize the interference from the cognate endogenous mRNAs. We obtained plasmid vectors for the two genes (IGF and PDGFRL) and conducted in vitro transcription (IVT). We have generated

SN143	SN186
A	1.1
B	1.7
C	1.8
D	1.9
E	1.10
F	1.12

Multi Dimensional Scaling (MDS) Analysis

Figure 4. Transcriptomics analysis of tumor subpopulations derived from two glioblastoma patients.

two IVT mRNAs, and have validated their expressions by regular qPCR Taqman assays (Figure 6). We are in the process of assessing whether or not these mRNAs will work on the Nanostring platform. We are testing whether the primer sets for IGF and PDGFRL used for pre-amplification in the Nanostring protocol are efficient in the pool of primers along with our other targets.
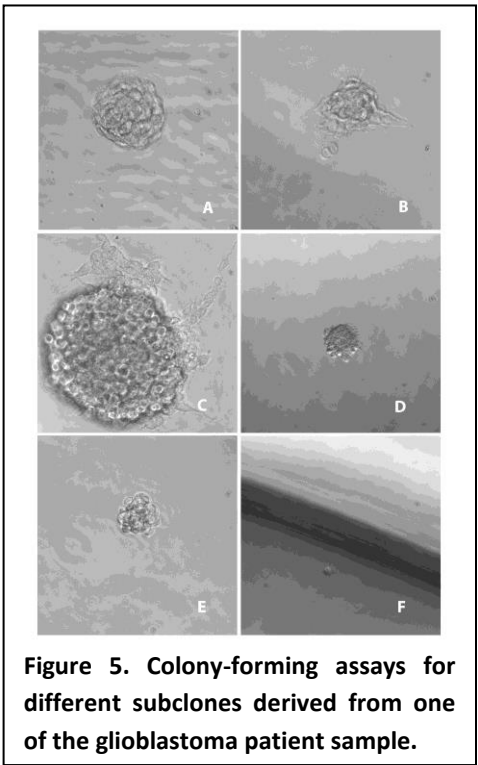

Figure 5. Colony-forming assays for different subclones derived from one of the glioblastoma patient sample.

We are now expanding the target gene panel (up to 500 genes to include key cancer signaling molecules, transcription factors, in addition to a panel of cell surface CD markers), as well as the cell numbers to be analyzed in 5 patient samples. We have performed these sets of experiments to enable new technology to be explored that can provide higher sensitivity, thereby reducing noise, for the detection of a set of stem cell markers and other genes suitable to determine the differentiation of cell populations. Both SN143 and SN186 are excellent samples to develop the cell line establishment, quantized cell population generation and transcriptome analysis experimentation to refine the technology, provide established proof of quantized cell populations and demonstrate the sensitive and

quantitative analysis of ultra-sensitive transcriptome profiling however, these two patients are not suitable for family sequencing. As described in our original Aim 1. From these combined analysis, we have been able to establish a minimal panel of 48 genes as potential glioblastoma subpopulation-specific markers. This panel is being incorporated into the complete panel of markers to be evaluated for targeted proteomics assays by selected reaction monitoring (SRM) of both cell line secretome and patient blood samples for biomarker development.

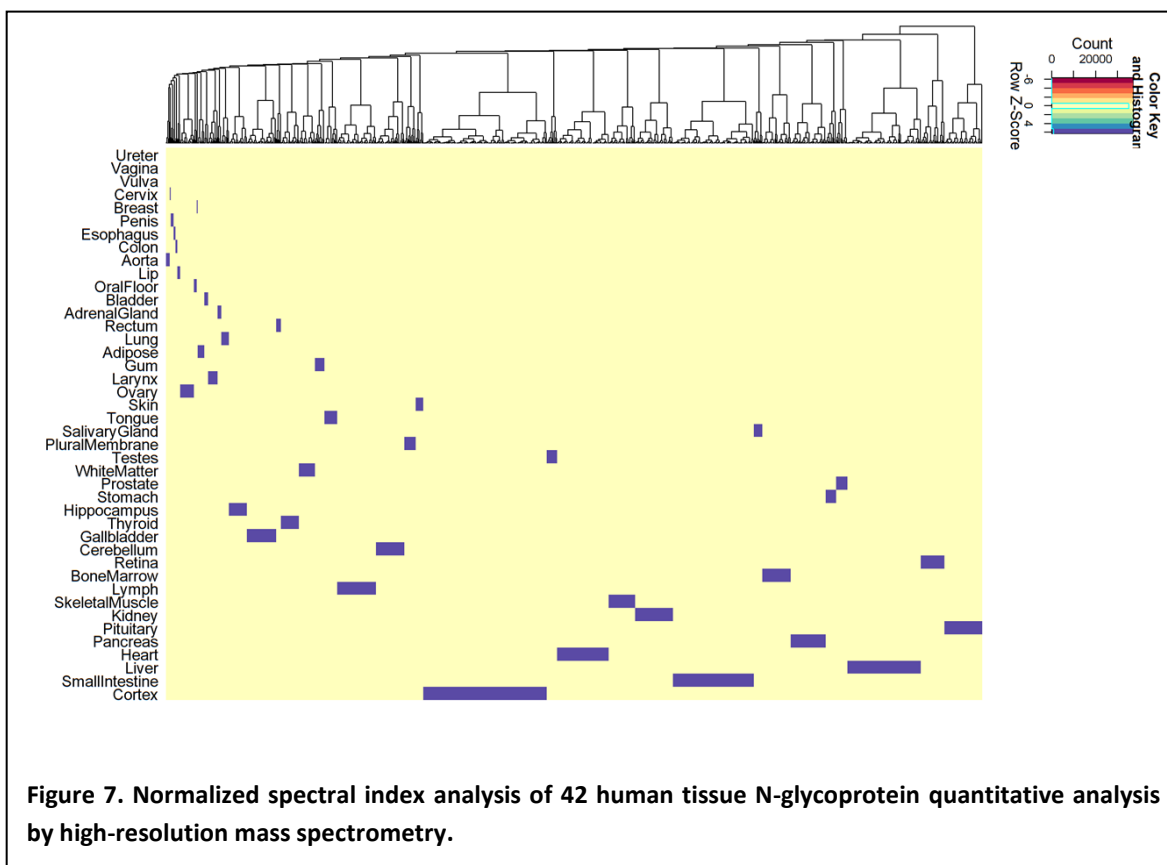**Figure 6. Quantitative real-time PCR Taqman assay of *in-vitro* transcribed IGF mRNAs.**
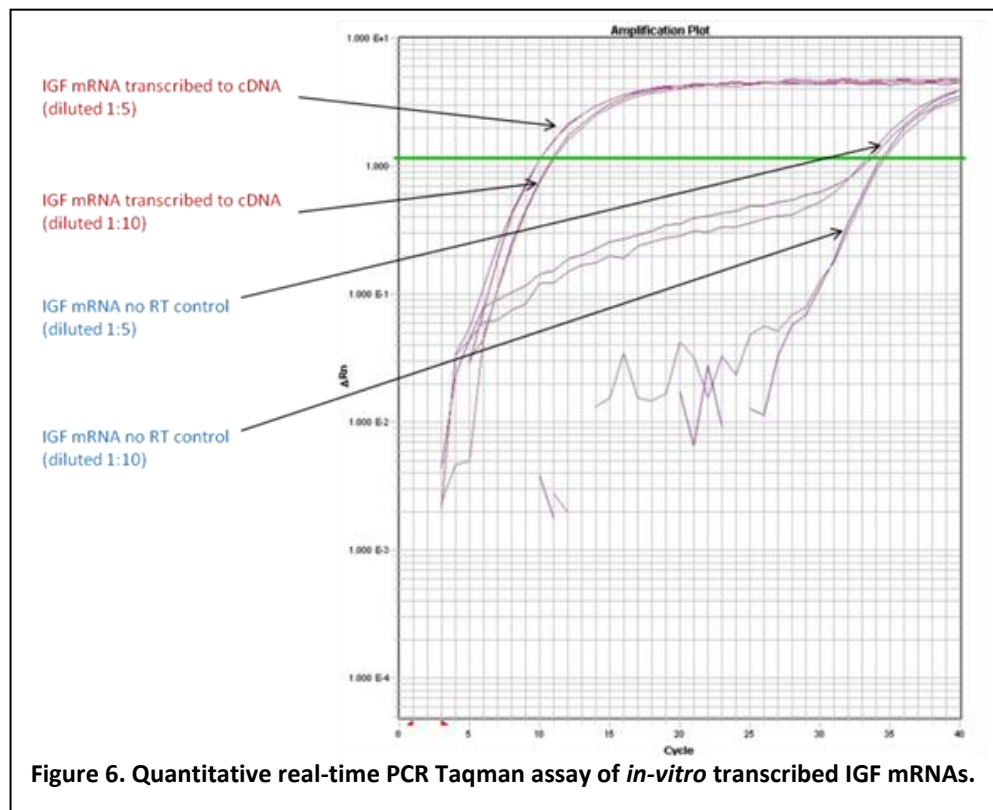
Experience gained from this exercise is being applied to the establishment of clonal cultures from the patient whose family members also consented to contribute samples. To this end, a third cell line (SN291) for which quantized cell populations are currently being established was obtained from a patient that has consented to provide blood samples from the progeny of this patient to allow family genome sequencing as stated in our original Aim 1. Blood was collected from this patient (SN291), as well as one sibling, and two children from this patient. The Ivy center is actively recruiting additional family-based consenting patients to complete the cohort of 5 patients for quantized cell creation and analysis. We anticipate in the coming few months we will be able to generate enough materials from this patient family for WGS and proteomics analysis and we expect to finish the analysis of all 5 patients tumor by December 2013.

Aim 3. For biomarker assessment by proteomics analysis, we have further refined a set of brain

specific proteins that can distinguish this tissue from other human tissues. Our targeted selected reaction monitoring (SRM) assays for these are now well established and will be used as standard control markers. These assays will also be released to the public for other researchers to use from the Moritz group Human SRMAtlas. We have selected 48 markers to coincide with the transcript markers selected for the single cell analysis that includes proteins such as the PDGF receptor and Kinesin family members and have established these as assays for the quantized cell lines and their subclones from the GBM patients.

One of the largest components of the proteomics assays will be determining the protein concentrations of the selected genes identified in the transcript analysis at both the quantized cell level and at the levels present in patients whole tumor samples and ultimately in blood plasma. Our studies on protein concentration with limiting amounts of total cell count have been encouraging and we continue to refine the techniques of protein extraction and analysis. For our test experiments, we limited the cell count to 100,000 cells and defined protein extraction protocols to determine the total protein weight that could be achieved. Our experiments showed that we could increase our total protein extraction almost 3 fold by using a combination of mass spectrometry compatible detergents as well as varying buffer combinations. In addition to these experiments, we have established a secretome analysis of several of these quantized cell lines to determine abundant expressed and secreted protein products from these cells as potential marker by combining this data with our established brain tissue secretome we have established. Differentially expressed proteins will be identified as high-value targets as these can uniquely identify glioblastoma cell from normal brain tissue. Our protein expression data is derived from our proteomic analysis of normal tissue called the N-glyco body map which comprises of cell surface and expressed glycoproteins determined to be differentially expressed amongst 42 distinct human tissues from around the body (Figure 7). This data has been compiled and will be published in 2013. Tissue specific proteins expressed and released into the blood from these tissues show and average of 4 proteins per tissue to be unique and will be incorporated into our overall proteomic analysis of glioblastoma quantized cells. We will incorporate these protein extraction conditions into our assay development pipeline and also produce SRM assays to detect and quantitate these proteins in the cell derived proteomes and patients serum as GBM signature markers. In addition, as part of a large NIH ARRA funded program, we have now completed creating multiple assays for essentially every human protein. With this tremendous resource, we will be easily able to select assays to detect the identified proteins in patient samples and we develop signatures from the genomic analysis of quantized cell populations form human glioblastoma patients. These signatures have the potential to be early indicators of disease and will be tremendously valuable to detect the disease early in future patient cohorts.

Figure 7. Normalized spectral index analysis of 42 human tissue N-glycoprotein quantitative analysis by high-resolution mass spectrometry.

We have not encountered technical problems and at this stage that we could not solve and we do not anticipate problems with the proposed work schedule for the next 12 months. Of course, in consideration of availability of human tumor samples, there may be additional small delays due to the availability of patients being seen by our collaborative group (CA100459P1, Award Number W81XWH-11-1-0488, Swedish Health Services), but there are continuously new patients that are being seen regularly as we expected and we assume this will continue at the current rate of patients presenting to our collaborators.

Our efforts in whole genome sequencing have advanced since the proposal and bode well for the upcoming work to be completed. Our efforts in the development of whole genome sequencing from families have been published [1] and our efforts in the development of quantitative assays for essentially every human protein have been successfully established. This program will be enabled through these significant developments and will provide the technical capability to provide significant results over the program in the next 12 months.

KEY RESEARCH ACCOMPLISHMENTS

- 2 cell lines for which quantized cell populations have been established (SN143 and SN186).
- A third cell line (SN291) for which there is full family consent and both blood and tumor was collected from this patient, blood from one sibling, and blood form the two children of this patient was collected and stored with procedures that are suitable for both cellular analysis, genomic analysis and proteomic analysis of these patients. Quantized cell populations are currently being established.
- Cell culture conditions suitable for propagation of quantized cells have been established.
- Transcriptomic profiling of hundreds of single cells from quantized populations of GBM tumors. Expression patterns of selected genes among these single tumor cells that will digitally stratify tumor cells into distinct populations have been established.
- Established SRM assays for 48 genes identified for transcriptome analysis of quantized cells for the measurement of protein levels in lysates of separate quantized cell populations.
- Developed cell culture conditions for secretome analysis of quantized cells and protein extraction conditions to maximize the amount of protein for high-mass accuracy quantitative mass spectrometry.
- Developed tissue specific cell surface N-glyco body map of 42 human tissues from around the body to define tissue specific proteins unique to each distinct tissues. This will allow the comparative analysis of differentially expressed proteins identified in the secretome of GBM quantized cell populations.

REPORTABLE OUTCOMES

No reportable outcomes have been established for 2011/2012/2013 period

CONCLUSION

Description of work to be performed during the next reporting period.

Over the next 12 months starting in July 2013 to June 2014 we will concentrate on the following aims:

In the first aim, we will complete our establishment of isolating up to 1000 cells from each of five human glioblastomas and quantify initially 500 different transcripts from each cell

(transcription factors, CD molecules, relevant signal transduction pathways, etc).

GBM patient sample collection. Through the collaboration with the Ivy Center for Advanced Brain Tumor Treatment at Swedish Neuroscience Institute in Seattle, we will access freshly excised GBM tumor samples for establishment of the quantized cell populations for further analysis Our initial analysis has included 3 separate GBM patients which has included one family with full consent for members to provide blood for whole genome sequencing and proteomic analysis.

We will complete our selection of 64 patients for proteomic analysis, maximally leveraging data generated from the TCGA and Allen atlas projects. Promising candidate target proteins will be validated on all 64 patient tumor and blood samples during this year by using targeted proteomic SRM approaches. We finalize this list by adding 36 control patients to bring our sample population to 100.

As we establish the final major quantized glioblastoma cells, we will complete aim 4 as proposed by selecting 10 to 20 cells from each major quantized glioblastoma cell type from two patients to be used to determine the complete genome sequences. We will also determine the normal genome sequences of each patient and their family members to enable the Mendelian-based error correction process recently described in our recently published Science paper [1]. The mutations will be analyzed against quantitative changes in the transcriptomes, miRNAomes and proteomes and against the relevant biological networks. We will also select signatures for combining into a target list for subsequent proteomics efforts. This aim will be completed by 2013 or earlier dependent on the success of the establishment of quantized glioblastoma cells from the final consented family.

REFERENCES

1.   J. C. Roach *et al.*, Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**, 636 (Apr 30).

APPENDICES

N/A